Submitted to the AHIMA workshop on "Solutions to Accelerate SNOMED CT Implementation in EHR Systems"
draft material -- please do not cite

Practical issues in mapping SNOMED CT neoplasms to ICD-O

Olivier Bodenreider[1], Anita Burgun[2]
[1]U.S. National Library of Medicine, NIH, Bethesda, Maryland, USA
[2]EA 3888, School of Medicine, IFR 140, Rennes University, Avenue Pr Léon Bernard, 35043 Rennes Cedex, France

[1]olivier@nlm.nih.gov, [2]anita.burgun@univ-rennes1.fr

Motivation: External reporting of clinical information for administrative or research purposes usually requires the information to be coded to a specific code set (e.g., ICD-9-CM, MedDRA, ICD-O). However, because most electronic health records rely on fine-grained clinical terminologies, such as SNOMED CT, to encode this information, there is a need for an automatic mapping from clinical terminologies to these specific code sets. Unlike the mapping developed to ICD-9-CM, no mapping is freely available between SNOMED CT and the International Classification of Diseases for Oncology (ICD-O), which represents an impediment for reporting clinical information to cancer registries. The objective of this paper is to explore the automatic mapping SNOMED CT neoplasms to ICD-O.

Background: The two major dimensions in the description of neoplasms are topography and morphology. ICD-O and SNOMED have a long history of sharing morphology codes, making ICD-O and SNOMED CT virtually compatible for this dimension. While topography codes are distinct in ICD-O and SNOMED CT, both terminologies are integrated in the NCI Metathesaurus. As a result, ICD-O and SNOMED CT codes for a given topography are identified as representing the same entity (i.e., as the same concept) in the NCI Metathesaurus. Our hypothesis is that, through shared morphology codes on the one hand and through shared topography concepts in the NCI Metathesaurus on the other, it should be easy to create a mapping to ICD-O for the 6,217 neoplasm concepts in SNOMED CT. The version of the NCI Metathesaurus used in this study is 2006AD, which contains the version of SNOMED CT dated of July 31, 2005 and ICD-O3.

Methods: For each neoplasm concept from SNOMED CT, we extract the related topography concepts (role: Finding site) and morphology concepts (Associated morphology). Morphology concepts are mapped to ICD-O through the SNOMEDID field of the SNOMED CT concept, which simply needs to be reformatted (e.g., the SNOMEDID M-97273 becomes 9727/3 in ICD-O). Topography codes are mapped through a common concept in the NCI Metathesaurus, when ICD-O and SNOMED CT codes for a given topography are linked to the same NCI Metathesaurus concept.

Conclusions: Quite unexpectedly, only 48% of the 6,217 SNOMED CT neoplasms can be mapped automatically to ICD-O (with at least one topography code and one morphology code). Overall, the mapping is generally more successful for morphology codes (65%) than for topography codes (54%). Of note, only 63% of the 409 topography codes and 34% of the 1,091 morphology codes from ICD-O are mapped to using the proposed method. As a consequence, the information originally coded to SNOMED CT and subsequently mapped to ICD-O might differ from the information directly coded to ICD-O. These differences are illustrated with examples, analyzed and discussed.